Joint Detection and Decoding in the Presence of Prior Information With Uncertainty

Suat Bayram, *Member, IEEE*, Berkan Dulek, *Member, IEEE*, and Sinan Gezici, *Senior Member, IEEE*

Abstract—**An optimal decision framework is proposed for joint detection and decoding when the prior information is available with some uncertainty. The proposed framework provides tradeoffs between the average inclusive error probability (computed using estimated prior probabilities) and the worst case inclusive error probability according to the amount of uncertainty while satisfying constraints on the probability of false alarm and the maximum probability of miss-detection. Theoretical results that characterize the structure of the optimal decision rule according to the proposed criterion are obtained. The proposed decision rule reduces to some well-known detectors in the case of perfect prior information or when the constraints on the probabilities of miss-detection and false alarm are relaxed. Numerical examples are provided to illustrate the theoretical results.**

Index Terms—**Bayes, decoding, detection, Neyman–Pearson.**

I. INTRODUCTION

IN MOST problems pertaining to hypothesis testing, one needs to perform a specific task depending on the chosen hypothesis. For example, in a sparse communication scenario, where an asynchronous receiver frequently observes pure noise due to a silent transmitter or the communication is jeopardized by the presence of a jammer, it is required that the receiver should reliably detect the presence of a message signal before attempting to decode it. In a wireless communications scenario, the channel characteristics may change abruptly due to blockage by a large obstacle or interference from other users, which in turn necessitates the detection of such changes before performing decoding. While the traditional approach has been to separately optimize for different tasks, joint optimization often provides improved performance as demonstrated for various frameworks such as joint detection and estimation [1]–[3], joint detection and source coding [4], and most recently, for joint detection and decoding [5], [6].

In this letter, we build upon the work conducted in [5] for the problem of joint detection and decoding over sparse communication channels. In that framework, the task was to detect whether a signal is emitted by the transmitter, and if so, to de-

code the message. The authors considered three figures of merit to measure performance:

1) the probability of false alarm (FA)—that is, the probability of deciding that some symbol was transmitted when in fact, the transmitter was silent meaning that pure noise was received;
2) the probability of miss-detection (MD)—that is, the probability of deciding that no transmission occurred when actually, the transmitter sent some message; and
3) the probability of inclusive error (IE)—that is, the probability of not correctly deciding on the actually transmitted symbol given that some message is transmitted, which includes both events of erroneous decoding and MD.

The optimum decision rule that minimizes the IE probability subject to constraints on the FA and the MD probabilities was derived.

Within the context of communication theory, the prior probabilities that are employed in the calculation of the IE and MD probabilities are usually assumed to be equally likely. However, in a more general decision theoretic setting, distinct signals that need to be decoded may assume unequal prior probabilities depending on the relative frequency of the events they symbolize. For example, in many practical image and speech compression applications, it is more realistic to model the output bit stream from the source encoder as an independent and identically distributed *nonuniform* Bernoulli process due to the suboptimality of the compression scheme [7]. Furthermore, the prior probabilities may not even be known precisely. In such circumstances, a sensible choice is to estimate the prior probabilities with some uncertainty. This estimate can then be employed in order to specify the average IE and MD probabilities but the uncertainty also needs to be controlled by restricting the worst case values for the corresponding probabilities in an analogous manner to the restricted Bayesian framework [8], [9]. To address these issues for the problem of joint detection and decoding, an optimal decision framework that takes into account the uncertainty in the prior information is proposed for the first time in this letter. Another advantage of the proposed framework is that it facilitates tradeoffs among three well-known decision criteria: Bayesian, minimax, and Neyman–Pearson since all three can be obtained as special cases.

II. PROBLEM FORMULATION AND THEORETICAL RESULTS

Following [5] and [6], we consider a hypothesis testing problem in which the probability law for the observed random variable Y belongs to one of the two disjoint sets. Only messages that belong to one of the sets are decoded whereas no decoding takes place if a decision is declared in favor of the other set. Such a scenario may arise in, for example, block coded communications, where in each block, the transmitter is either silent or transmits a codeword from a given codebook. The task of the detector is then to decide whether a codeword is transmitted (or pure noise is observed at the receiver), and if so, to decode

it (see [6] for various motivations). In this framework, the null and alternative hypotheses can be represented, respectively, as

$$\mathcal{H}_0 : Y \sim p_0(y), \quad \mathcal{H}_i : Y \sim p_i(y) \text{ for } i = 1, \ldots, M \quad (1)$$

where $p_i(y)$ denotes the probability density function (pdf) of observation Y under hypothesis $\mathcal{H}_i$ for $i \in \{0, 1, \ldots, M\}$.[1] It is assumed that there exists some prior information, albeit uncertain, related to the alternative hypotheses. In particular, given that the null hypothesis is false (e.g., a codeword is transmitted), the probability that the alternative hypothesis $\mathcal{H}_i$ is true (e.g., the ith codeword is transmitted) is denoted as π_i such that $\sum_{i=1}^{M} \pi_i = 1$ and $\pi_i \geq 0 \ \forall i \in \{1, \ldots, M\}$. For example, in the case of block coded communications [5], all codewords are equiprobable *a priori*, i.e., $\pi_i = 1/M \ \forall i \in \{1, \ldots, M\}$, where M denotes the number of distinct codewords in the codebook. Let $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_M)$. The observation set is denoted with Γ and the decision rule ϕ is a partition of Γ into $M + 1$ regions, denoted by $\{\Gamma_i\}_{i=0}^{M}$. If $y \in \Gamma_i$ for some $1 \leq i \leq M$, then the ith message represented by $\mathcal{H}_i$ is decoded. If $y \in \Gamma_0$ (i.e., the *rejection region*), then hypothesis $\mathcal{H}_0$ is identified and no decoding takes place.

The probability of selecting $\mathcal{H}_i$ when $\mathcal{H}_j$ is true is denoted by $P_j(\Gamma_i)$ and calculated as $P_j(\Gamma_i) = \int_{\Gamma_i} p_j(y) \, dy$. Consequently, the IE probability can be written as [5]

$$P_{\text{IE}}(\phi, \boldsymbol{\pi}) = \sum_{i=1}^{M} \pi_i P_i(\overline{\Gamma}_i) \quad (2)$$

where $\overline{\Gamma}_i$ denotes the complement of the set Γ_i, i.e., $\overline{\Gamma}_i = \Gamma \setminus \Gamma_i$, and the probability of correct decision (CD) is given by

$$P_{\text{CD}}(\phi, \boldsymbol{\pi}) = \sum_{i=1}^{M} \pi_i P_i(\Gamma_i) \quad (3)$$

which satisfy $P_{\text{IE}}(\phi, \boldsymbol{\pi}) + P_{\text{CD}}(\phi, \boldsymbol{\pi}) = 1$. In addition, the probabilities of MD and FA are expressed, respectively, as

$$P_{\text{MD}}(\phi, \boldsymbol{\pi}) = \sum_{i=1}^{M} \pi_i P_i(\Gamma_0). \quad (4)$$

$$P_{\text{FA}}(\phi) = P_0(\overline{\Gamma}_0). \quad (5)$$

In practice, the conditional prior probabilities (i.e., the elements of $\boldsymbol{\pi}$) are known with some uncertainty (e.g., estimated). Let $\boldsymbol{\pi}^{\text{est}}$ denote an estimate for the conditional prior probabilities. The proposed hypothesis testing framework is as follows:

$$\underset{\phi}{\text{minimize}} \ \lambda \, P_{\text{IE}}(\phi, \boldsymbol{\pi}^{\text{est}}) + (1 - \lambda) \max_{\boldsymbol{\pi}} P_{\text{IE}}(\phi, \boldsymbol{\pi}) \quad (6a)$$

$$\text{subject to} \max_{\boldsymbol{\pi}} \ P_{\text{MD}}(\phi, \boldsymbol{\pi}) \leq \beta, \quad (6b)$$

$$P_{\text{FA}}(\phi) \leq \alpha \quad (6c)$$

where $\lambda \in [0, 1]$ is a design parameter that is set according to the degree of uncertainty in the knowledge of $\boldsymbol{\pi}^{\text{est}}$. In particular, a larger value of λ is associated with less uncertainty in the conditional prior probabilities. According to this formulation, the objective function is a combination of the average IE probability (obtained based on $\boldsymbol{\pi}^{\text{est}}$) and the worst case IE probability, and the worst case (i.e., maximum) MD probability and the FA probability are not allowed to exceed predefined thresholds

(α and β, respectively).[2] The formulation proposed in (6) generalizes that given in [5, eq. (10)] in the sense that the uncertainty in the conditional prior probabilities is also taken into account via (6a) and (6b).

To characterize the optimal detector corresponding to (6), a two-step approach is taken in the following: (i) The optimal decision regions for the alternative hypotheses are obtained for a given decision region for the null hypothesis in Lemma 1. (ii) The form of the optimal detector is specified completely in Proposition 1 based on the result in Lemma 1.

Lemma 1: Consider the optimization problem in (6) when the decision region Γ_0 for the null hypothesis is fixed (given) and suppose that the MD and FA constraints in (6) are satisfied for the given set Γ_0. Then, the optimal decision regions $\{\Gamma_i\}_{i=1}^{M}$ corresponding to the solution of (6) are obtained as

$$\Gamma_i = \left\{ y \in \overline{\Gamma}_0 : \pi_i^{\text{ls}} p_i(y) \geq \pi_k^{\text{ls}} p_k(y) \quad \forall k \neq i, k \neq 0 \right\} \quad (7)$$

for all $i \in \{1, \ldots, M\}$ with $\boldsymbol{\pi}^{\text{ls}} = (\pi_1^{\text{ls}}, \ldots, \pi_M^{\text{ls}})$ denoting the least favorable conditional prior distribution given by

$$\boldsymbol{\pi}^{\text{ls}} = \lambda \, \boldsymbol{\pi}^{\text{est}} + (1 - \lambda) \, \boldsymbol{\pi}^{\mu} \quad (8)$$

where $\boldsymbol{\pi}^{\mu} = (\pi_1^{\mu}, \ldots, \pi_M^{\mu})$ satisfies the following relation[3]:

$$\sum_{i=1}^{M} \pi_i^{\mu} P_i(\overline{\Gamma}_i) = \max \left\{ P_1(\overline{\Gamma}_1), \ldots, P_M(\overline{\Gamma}_M) \right\}. \quad (9)$$

Proof: From (4)–(6), the FA probability and the maximum MD probability are expressed as $P_{\text{FA}}(\phi) = 1 - \int_{\Gamma_0} p_0(y) dy$ and $\max_{\boldsymbol{\pi}} P_{\text{MD}}(\phi, \boldsymbol{\pi}) = \max_{\boldsymbol{\pi}} \sum_{i=1}^{M} \pi_i \int_{\Gamma_0} p_i(y) dy = \max \{ \int_{\Gamma_0} p_1(y) dy, \ldots, \int_{\Gamma_0} p_M(y) dy \}$. Hence, when the decision region Γ_0 for the null hypothesis is given, the FA probability and the maximum MD probability are fixed since they only depend on Γ_0 and the pdfs under different hypotheses, namely $\{p_i(y)\}_{i=0}^{M}$. Since the maximum MD and the FA constraints in (6) are assumed to be satisfied for given Γ_0, the problem in (6) reduces to

$$\underset{\{\Gamma_i\}_{i=1}^{M}}{\text{minimize}} \ \lambda \, P_{\text{IE}}(\phi, \boldsymbol{\pi}^{\text{est}}) + (1 - \lambda) \max_{\boldsymbol{\pi}} P_{\text{IE}}(\phi, \boldsymbol{\pi}). \quad (10)$$

Using (2), the optimization in (10) can be written as

$$\underset{\{\Gamma_i\}_{i=1}^{M}}{\text{minimize}} \ \lambda \sum_{i=1}^{M} \pi_i^{\text{est}} P_i(\overline{\Gamma}_i)$$

$$+ (1 - \lambda) \max \left\{ P_1(\overline{\Gamma}_1), \ldots, P_M(\overline{\Gamma}_M) \right\}. \quad (11)$$

The formulation in (11) is known as the *restricted(-risk) Bayesian* problem [8], [10], which aims at finding the decision regions $\{\Gamma_i\}_{i=1}^{M}$ that minimize the Bayes risk (computed using the estimated/uncertain conditional prior probabilities) subject to a constraint, induced by λ, on the worst case conditional risk (see [10, eq. 2.2]). The solution of the restricted Bayesian problem is given by the Bayes rule corresponding to the *least favorable* distribution, which is specified by a mixture of the estimated conditional prior distribution with another conditional prior distribution as in (8) [8], [10]. In particular, the Bayes rule corresponding to $\boldsymbol{\pi}^{\text{ls}}$ in (8) yields a solution of (11) if the

---

[1] In an analogy with the block coded communications scenario, $\mathcal{H}_0$ represents pure noise reception while $\mathcal{H}_i$ indicates that a noise corrupted version of the ith codeword is received.

[2] For a given α, the minimum value of β is set by the solution of the max–min Neyman–Pearson problem [12]. Hence, for a given α, the value of β must be selected to be greater than that minimum value.

[3] The relation in (9) expresses the intrinsic equalizer nature of minimax problems (see [8], [10]).

condition in (9) holds [10, Th. 1]. Finally, since the solution of (11) [equivalently, (10)] is the Bayes rule corresponding to $\boldsymbol{\pi}^{\mathrm{ls}}$, it is given by the maximum *a posteriori* probability decision rule (under uniform cost assignment [11]), that is, a decision is declared in favor of the ith hypothesis if $\pi_i^{\mathrm{ls}} p_i(y)$ is larger than or equal to $\pi_j^{\mathrm{ls}} p_j(y)$ for all $j \in \{1, \ldots, M\} \setminus \{i\}$, where $i \in \{1, \ldots, M\}$. Hence, the optimal decision regions are obtained as in (7).

Lemma 1 states that once Γ_0 is determined, the remaining decision regions $\{\Gamma_i\}_{i=1}^M$ are optimally specified via (7) in conjunction with (8) and (9). Therefore, we focus on obtaining the optimal decision rule for the rejection region Γ_0 in the remainder of the manuscript. The following proposition characterizes the optimal rejection region corresponding to the null hypothesis, which in turn yields the form of the solution to the optimization problem proposed in (6).

Proposition 1: Consider a decision rule ϕ^* specified by a partition of the observation space Γ into $M+1$ regions denoted as $\{\Gamma_i^*\}_{i=0}^M$ such that

$$\Gamma_0^* = \left\{ y \in \Gamma : a \sum_{i=1}^M \pi_i^{\mathrm{md}} p_i(y) + \max_{k \neq 0} \widetilde{\pi}_k^{\mathrm{ls}} p_k(y) \leq b \, p_0(y) \right\} \tag{12}$$

for some nonnegative thresholds a and b that are adjusted to satisfy the constraints in (6b) and (6c) (see [5, discussions after the proof of Lemma 1]), $\{\Gamma_i^*\}_{i=1}^M$ are obtained from Γ_0^* as described in Lemma 1, $\widetilde{\boldsymbol{\pi}}^{\mathrm{ls}}$ is the corresponding least favorable distribution as specified in (8) and (9), and $\boldsymbol{\pi}^{\mathrm{md}}$ satisfies[4]

$$P_{\mathrm{MD}}(\phi^*, \boldsymbol{\pi}^{\mathrm{md}}) = \max_{i \in \{1, \ldots, M\}} P_i(\Gamma_0^*). \tag{13}$$

Then, for any other decision rule ϕ represented by $\{\Gamma_i\}_{i=0}^M$ that satisfies $\max_{\boldsymbol{\pi}} P_{\mathrm{MD}}(\phi, \boldsymbol{\pi}) \leq \max_{\boldsymbol{\pi}} P_{\mathrm{MD}}(\phi^*, \boldsymbol{\pi})$ and $P_{\mathrm{FA}}(\phi) \leq P_{\mathrm{FA}}(\phi^*)$, the following inequality holds:

$$\lambda \, P_{\mathrm{IE}}(\phi^*, \boldsymbol{\pi}^{\mathrm{est}}) + (1 - \lambda) \max_{\boldsymbol{\pi}} P_{\mathrm{IE}}(\phi^*, \boldsymbol{\pi})$$
$$\leq \lambda \, P_{\mathrm{IE}}(\phi, \boldsymbol{\pi}^{\mathrm{est}}) + (1 - \lambda) \max_{\boldsymbol{\pi}} P_{\mathrm{IE}}(\phi, \boldsymbol{\pi}). \tag{14}$$

That is, the optimal solution to (6) is characterized by the rule ϕ^* with decision regions $\{\Gamma_i^*\}_{i=0}^M$.

Proof: From (12), it is seen that the following inequality holds for all $y \in \Gamma$:

$$(\mathcal{I}\{y \in \Gamma_0^*\} - \mathcal{I}\{y \in \Gamma_0\})$$
$$\times \left(b \, p_0(y) - a \sum_{i=1}^M \pi_i^{\mathrm{md}} p_i(y) - \max_{k \neq 0} \widetilde{\pi}_k^{\mathrm{ls}} p_k(y) \right) \geq 0 \tag{15}$$

where $\mathcal{I}\{\cdot\}$ denotes the indicator function.[5] Since the expression in (15) is nonnegative $\forall y \in \Gamma$, its integral over the observation set Γ is also nonnegative, which leads to the following inequality:

$$b \left(P_{\mathrm{FA}}(\phi) - P_{\mathrm{FA}}(\phi^*) \right) + a \left(P_{\mathrm{MD}}(\phi, \boldsymbol{\pi}^{\mathrm{md}}) - P_{\mathrm{MD}}(\phi^*, \boldsymbol{\pi}^{\mathrm{md}}) \right)$$
$$\geq \int_{\Gamma_0^*} \max_{k \in \{1, \ldots, M\}} \widetilde{\pi}_k^{\mathrm{ls}} p_k(y) \, dy - \int_{\Gamma_0} \max_{k \in \{1, \ldots, M\}} \widetilde{\pi}_k^{\mathrm{ls}} p_k(y) \, dy \tag{16}$$

---

[4]The relation in (13) is due to the intrinsic equalizer nature of the max–min Neyman–Pearson problem (see [12] and references therein).
[5]This proof approach is similar to that of the Neyman–Pearson lemma [11, p. 24] and that in [5, Lemma 1].

where $P_{\mathrm{FA}}(\phi) = P_0(\overline{\Gamma}_0) = 1 - P_0(\Gamma_0)$ and $P_{\mathrm{MD}}(\phi, \boldsymbol{\pi}^{\mathrm{md}}) = \sum_{i=1}^M \pi_i^{\mathrm{md}} P_i(\Gamma_0)$ are substituted. Then, we get

$$P_{\mathrm{MD}}(\phi, \boldsymbol{\pi}^{\mathrm{md}}) \leq \max_{\boldsymbol{\pi}} P_{\mathrm{MD}}(\phi, \boldsymbol{\pi}) \leq \max_{\boldsymbol{\pi}} P_{\mathrm{MD}}(\phi^*, \boldsymbol{\pi})$$
$$= P_{\mathrm{MD}}(\phi^*, \boldsymbol{\pi}^{\mathrm{md}}) \tag{17}$$

where the first inequality is by definition, the second inequality $\max_{\boldsymbol{\pi}} P_{\mathrm{MD}}(\phi, \boldsymbol{\pi}) \leq \max_{\boldsymbol{\pi}} P_{\mathrm{MD}}(\phi^*, \boldsymbol{\pi})$ is assumed in the proposition, and the equality is due to (13). Using $P_{\mathrm{FA}}(\phi) \leq P_{\mathrm{FA}}(\phi^*)$ in conjunction with (17), the left-hand side of (16) is seen to be nonpositive, from which it follows that $\int_{\Gamma_0^*} \max_{k \in \{1, \ldots, M\}} \widetilde{\pi}_k^{\mathrm{ls}} p_k(y) \, dy \leq \int_{\Gamma_0} \max_{k \in \{1, \ldots, M\}} \widetilde{\pi}_k^{\mathrm{ls}} p_k(y) \, dy$. This relation implies that

$$\int_{\overline{\Gamma}_0^*} \max_{k \in \{1, \ldots, M\}} \widetilde{\pi}_k^{\mathrm{ls}} p_k(y) \, dy \geq \int_{\overline{\Gamma}_0} \max_{k \in \{1, \ldots, M\}} \widetilde{\pi}_k^{\mathrm{ls}} p_k(y) \, dy. \tag{18}$$

The integral on the right-hand side of (18) can be bounded as follows:

$$\int_{\overline{\Gamma}_0} \max_{k \in \{1, \ldots, M\}} \widetilde{\pi}_k^{\mathrm{ls}} p_k(y) \, dy = \sum_{i=1}^M \int_{\Gamma_i} \max_{k \in \{1, \ldots, M\}} \widetilde{\pi}_k^{\mathrm{ls}} p_k(y) \, dy$$
$$\geq \sum_{i=1}^M \int_{\Gamma_i} \widetilde{\pi}_i^{\mathrm{ls}} p_i(y) \, dy = \sum_{i=1}^M \widetilde{\pi}_i^{\mathrm{ls}} P_i(\Gamma_i). \tag{19}$$

On the other hand, based on (7) in Lemma 1, the integral on the left-hand side of (18) is recognized as the CD probability corresponding to observation set $\overline{\Gamma}_0^*$; i.e., $\sum_{i=1}^M \widetilde{\pi}_i^{\mathrm{ls}} P_i(\Gamma_i^*)$. Hence, from (18) and (19), the following relation is obtained: $\sum_{i=1}^M \widetilde{\pi}_i^{\mathrm{ls}} P_i(\Gamma_i^*) \geq \sum_{i=1}^M \widetilde{\pi}_i^{\mathrm{ls}} P_i(\Gamma_i)$. This inequality can also be stated as $\sum_{i=1}^M \widetilde{\pi}_i^{\mathrm{ls}} (1 - P_i(\overline{\Gamma}_i^*)) \geq \sum_{i=1}^M \widetilde{\pi}_i^{\mathrm{ls}} (1 - P_i(\overline{\Gamma}_i))$, which leads to the following expression after substituting (8):

$$\lambda \sum_{i=1}^M \pi_i^{\mathrm{est}} P_i(\overline{\Gamma}_i^*) + (1 - \lambda) \sum_{i=1}^M \widetilde{\pi}_i^{\mu} P_i(\overline{\Gamma}_i^*)$$
$$\leq \lambda \sum_{i=1}^M \pi_i^{\mathrm{est}} P_i(\overline{\Gamma}_i) + (1 - \lambda) \sum_{i=1}^M \widetilde{\pi}_i^{\mu} P_i(\overline{\Gamma}_i). \tag{20}$$

Since $\sum_{i=1}^M \widetilde{\pi}_i^{\mu} P_i(\overline{\Gamma}_i^*) = \max_{\boldsymbol{\pi}} P_{\mathrm{IE}}(\phi^*, \boldsymbol{\pi})$ due to (9) and $\sum_{i=1}^M \widetilde{\pi}_i^{\mu} P_i(\overline{\Gamma}_i) = P_{\mathrm{IE}}(\phi, \widetilde{\boldsymbol{\pi}}^{\mu}) \leq \max_{\boldsymbol{\pi}} P_{\mathrm{IE}}(\phi, \boldsymbol{\pi})$ by definition [see (2)], the relation in (20) implies (14). Overall, it is proved that any decision rule with FA and maximum MD probabilities not exceeding those of the proposed decision rule ϕ^*, respectively, cannot achieve a lower value for objective function than the proposed one. Hence, the optimal solution to (6) is characterized by the rule ϕ^* with decision regions $\{\Gamma_i^*\}_{i=0}^M$ as specified in the proposition.

Proposition 1 states that the optimal decision rule corresponding to (6) has a rejection region specified by (12) and the decoding regions are obtained from (7). In order to solve (6), the decision regions specified in (7) and (12) are used in place of the decision rule ϕ, and (9) and (13) are incorporated into (6) as the additional constraints. As a result, we have an optimization problem with a finite number of unknowns (which are a, b, $\boldsymbol{\pi}^{\mu}$, and $\boldsymbol{\pi}^{\mathrm{md}}$). In the simulations, we have used a global optimization tool in MATLAB (namely, the particle swarm optimization algorithm) to solve this problem. It is worth mentioning that while the threshold a is mainly related to satisfying the maximum MD constraint (see footnote 6), the threshold b is primarily selected to satisfy the FA constraint. It should
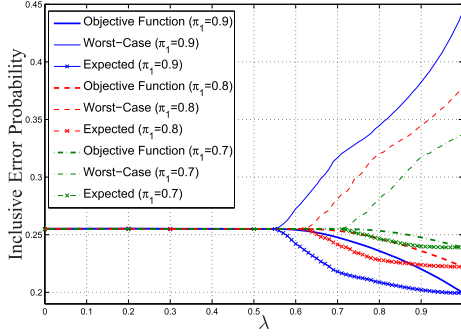
Fig. 1. Objective function in (6), $P_{\text{IE}}(\phi^*, \boldsymbol{\pi}^{\text{est}})$, and $\max_{\boldsymbol{\pi}} P_{\text{IE}}(\phi^*, \boldsymbol{\pi})$ versus λ for various values of the estimated conditional prior probabilities. A Gaussian mixture with $N_m = 4$ components is assumed, where the means are $\{-1.5, -0.5, 0.5, 1.5\}$ with corresponding weights $\{0.4, 0.1, 0.1, 0.4\}$ and the standard deviations of all the mixture components are $\sigma = 0.2$. The remaining parameters are set to $\alpha = 0.1$, $\beta = 0.22$, and $A = 1.3$.

be noted that for large values of a, the problem under consideration is essentially a Neyman–Pearson test between $p_0(y)$ and $\sum_{i=1}^{M} \pi_i^{\text{md}} p_i(y)$, known as the max–min Neyman–Pearson problem, where the aim is to minimize the worst case (i.e., maximum) MD probability under the FA constraint [12]. On the other hand, for small values of a, the problem converges to the restricted Bayesian problem, which covers both Bayesian and minimax frameworks as special cases, after proper normalization of the conditional prior probabilities along with a proper selection of b (i.e., $b = \widetilde{\pi}_0^{\text{ls}}$) [10]. More explicitly, the decision rule treats the null hypothesis as another message and decoding is performed among all hypotheses in the presence of prior uncertainty according to $\arg\max_{k \in \{0,1,\ldots,M\}} \widetilde{\pi}_k^{\text{ls}} p_k(y)$. Furthermore, in the case of perfect prior information, the formulation proposed in (6) reduces to that introduced in [5] when equal priors are assumed and the maximum MD constraint is replaced with the average MD constraint. As a result, our formulation can be considered as a generalization of Bayesian and Neyman–Pearson frameworks in the presence of uncertainty.

III. NUMERICAL RESULTS

In this section, the theoretical results in Section II are investigated via simulations. Consider a sensor that intermittently reports to a remote node about the temperature. There are idle periods during which no transmission occurs, indicating that the temperature is within acceptable limits. The sensor employs binary phase shift keying to signal when the temperature gets too hot or too cold. Consequently, at the receiver side, the following hypothesis testing problem arises: $\mathcal{H}_0 : Y = N$, $\mathcal{H}_1 : Y = A + N$, $\mathcal{H}_2 : Y = -A + N$, where N represents additive noise and $A > 0$ is a known signal level. Namely, when the transmitter is silent, the receiver observes pure noise. Otherwise, a constant signal level corrupted by additive noise is acquired. A symmetric Gaussian mixture noise with the following pdf is assumed [9]: $p_N(n) = \sum_{l=1}^{N_m} \xi_l \varphi_l(n - \mu_l)$, where N_m denotes the number of components in the mixture, μ_l is the mean value of the lth component, $\sum_{l=1}^{N_m} \xi_l = 1, \xi_l > 0$, and $\varphi_l(n) = (\sqrt{2\pi} \sigma_l)^{-1} \exp\{-n^2/(2\sigma_l^2)\}$ with σ_l representing the standard deviation of the lth mixture component for $l \in \{1, \ldots, N_m\}$. The mixture parameters $\{\xi_l, \mu_l, \sigma_l\}_{l=1}^{N_m}$ are selected to make the resulting pdf symmetric. It is assumed that the conditional prior probabilities of the hypotheses $\mathcal{H}_1$ and $\mathcal{H}_2$ corresponding to hot and cold states, respectively, are not
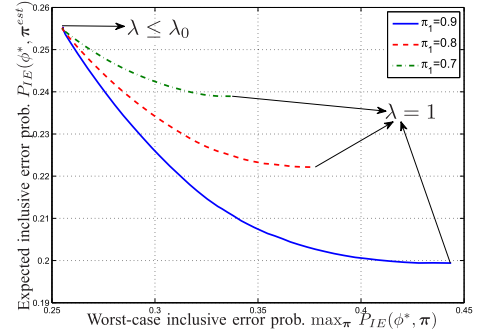

Fig. 2. Expected IE probability $P_{\text{IE}}(\phi^*, \boldsymbol{\pi}^{\text{est}})$ versus worst case IE probability $\max_{\boldsymbol{\pi}} P_{\text{IE}}(\phi^*, \boldsymbol{\pi})$ for the scenario in Fig. 1.

necessarily equal (see [7] for other examples of unequal priors), and they are estimated based on the past measurements. Due to suboptimality of the estimation procedure, there exists some degree of uncertainty in the conditional prior probabilities. Hence, employing the detection criterion introduced in (6) can be beneficial to compensate for the undesired effects of uncertainty on the system performance, as investigated below.

In Fig. 1, the objective function in (6) together with the expected and the worst case IE probabilities ($P_{\text{IE}}(\phi^*, \boldsymbol{\pi}^{\text{est}})$ and $\max_{\boldsymbol{\pi}} P_{\text{IE}}(\phi^*, \boldsymbol{\pi})$, respectively) are plotted against the change in λ for various values of the estimated conditional prior probabilities, where $\alpha = 0.1, \beta = 0.22$, and $A = 1.3$. When λ is less than or equal to a specific value (which is $\lambda_0 = 0.55$ for $\pi_1 = 0.9$, $\lambda_0 = 0.625$ for $\pi_1 = 0.8$, and $\lambda_0 = 0.714$ for $\pi_1 = 0.7$), the least favorable conditional prior distribution is realized as $\widetilde{\boldsymbol{\pi}}^{\text{ls}} = (0.5, 0.5)$ along with $\boldsymbol{\pi}^{\text{md}} = (0.5, 0.5)$, resulting in equal conditional IE probabilities (i.e., $P_1(\overline{\Gamma}_1) = P_2(\overline{\Gamma}_2)$) and equal conditional MD probabilities (i.e., $P_1(\Gamma_0) = P_2(\Gamma_0)$). This, in turn, results in equal expected and the worst case IE probabilities (that is, $P_{\text{IE}}(\phi^*, \boldsymbol{\pi}^{\text{est}}) = \max_{\boldsymbol{\pi}} P_{\text{IE}}(\phi^*, \boldsymbol{\pi})$), as observed in Fig. 1. Such behaviour can partly be attributed to the fact that small values of λ place more emphasis on the minimization of the worst case IE probability, which is realized with equal conditional priors for all $\lambda \leq \lambda_0$. The emphasis can be shifted towards the expected IE probability by considering higher values of λ (i.e., decreasing the amount of uncertainty in the prior information), which in turn renders lower scores for the objective function possible. For $\lambda > \lambda_0$, it is seen in Fig. 1 that the gap between the expected and the worst case IE probabilities widens as λ increases. In summary, the minimum value for $\max_{\boldsymbol{\pi}} P_{\text{IE}}(\phi, \boldsymbol{\pi})$ is achieved when $\lambda \leq \lambda_0$, whereas the minimum value for $P_{\text{IE}}(\phi, \boldsymbol{\pi}^{\text{est}})$ is obtained for $\lambda = 1$. Therefore, the proposed formulation in (6) provides the optimal tradeoff between the expected and the worst case IE probabilities (that is, $P_{\text{IE}}(\phi, \boldsymbol{\pi}^{\text{est}})$ and $\max_{\boldsymbol{\pi}} P_{\text{IE}}(\phi, \boldsymbol{\pi})$, respectively) subject to the constraints on the FA and the maximum MD probabilities by adjusting λ within the interval $[\lambda_0, 1]$ according to the uncertainty level. This is illustrated in Fig. 2.

IV. CONCLUDING REMARKS

For the joint detection and decoding problem, an optimal decision framework that takes into account the uncertainty in the prior information has been proposed. The proposed framework facilitates tradeoffs between the average IE probability and the worst case IE probability according to the amount of uncertainty while satisfying constraints on the probability of FA and the maximum probability of MD.

REFERENCES

[1] D. Middleton and R. Esposito, "Simultaneous optimum detection and estimation of signals in noise," *IEEE Trans. Inf. Theory*, vol. 14, no. 3, pp. 434–444, May 1968.

[2] B. Baygun and A. O. Hero, "Optimal simultaneous detection and estimation under a false alarm constraint," *IEEE Trans. Inf. Theory*, vol. 41, no. 3, pp. 688–703, May 1995.

[3] G. V. Moustakides, G. H. Jajamovich, A. Tajer, and X. Wang, "Joint detection and estimation: Optimum tests and applications," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4215–4229, Jul. 2012.

[4] N. Merhav, "Asymptotically optimal decision rules for joint detection and source coding," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 6787–6795, Nov. 2014.

[5] N. Weinberger and N. Merhav, "Codeword or noise? Exact random coding exponents for joint detection and decoding," *IEEE Trans. Inf. Theory*, vol. 60, no. 9, pp. 5077–5094, Sep. 2014.

[6] N. Weinberger and N. Merhav, "Channel detection in coded communication," *CoRR*, 2015. [Online]. Available: http://arxiv.org/abs/1509.01806

[7] H. Kuai, F. Alajaji, and G. Takahara, "Tight error bounds for nonuniform signaling over AWGN channels," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2712–2718, Nov. 2000.

[8] E. L. Lehmann, *Testing Statistical Hypotheses*, 2nd ed. New York, NY, USA: Chapman & Hall, 1986.

[9] S. Bayram, S. Gezici, and H. V. Poor, "Noise enhanced hypothesis-testing in the restricted Bayesian framework," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 3972–3989, Aug. 2010.

[10] J. L. Hodges, Jr., and E. L. Lehmann, "The use of previous experience in reaching statistical decisions," *Ann. Math. Stat.*, vol. 23, no. 3, pp. 396–407, Sep. 1952.

[11] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York, NU, USA: Springer-Verlag, 1994.

[12] S. Bayram and S. Gezici, "On the restricted Neyman-Pearson approach for composite hypothesis-testing in the presence of prior distribution uncertainty," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 5056–5065, Oct. 2011.